重大突发事件中社会群体信息行为舆情主题演化图谱构建研究*

王晰巍[1,2,3]，李聪聪[1]，孙哲[1]，李玥琪[1]
（1. 吉林大学商学与管理学院，吉林 长春 130015）
（2. 吉林大学大数据管理研究中心，吉林 长春 130015）
（3. 吉林大学网络空间治理研究中心，吉林 长春 130015）

摘 要：[目的/意义] 研究重大突发事件中社会群体信息行为舆情主题图谱构建，有利于更好地发现重大突发事件下舆情演进态势和识别敏感舆情话题，从而进行正向的舆情引导，对重大突发事件的应急和维护社会稳定具有重要作用。[方法/过程] 本文基于 LDA 构建社会群体信息行为舆情主题聚类演化图谱，基于时序性构建舆情主题热度演化图谱，基于相似度构建舆情主题路径演化图谱，并给出重大突发事件中社会群体信息行为舆情主题图谱构建过程模型。[结果/结论] 研究结果表明，利用本文构建的重大突发事件中社会群体信息行为舆情主题分析模型，可以对社交媒体中群体信息行为舆情主题特征、主题热度影响因素及主题最优传播路径进行分析。本文的研究为重大突发事件中社会群体信息行为舆情分析提供了新的理论框架和分析方法，为重大突发事件中的网络舆情引导和舆情治理提供参考。

关键词：重大突发事件；社交网络舆情；主题演化；群体信息行为
分类号：G25 G350

1 引言

2022 年 10 月中国共产党召开第二十次全国代表大会，习近平总书记在报告中指出“要完善公共安全体系，提高防灾减灾救灾和急难险重突发公共事件处置保障能力”，多次强调“坚持安全第一、预防为主，完善公共安全体系和国际应急管理体系建设”，为总体国家安全观下重大突发事件的应急管理工作指明了方向[1]。重大突发事件下由于信息的不对称，极易引发一系列重要信息的错误解读，甚至是不实言论的恶意传播，从而影响社会稳定，造成人民恐慌。同时，新媒体和社交网络是国家机关发布相关政策的重要渠道和人民群众日常交流信息的重要平台，对网络社会群体信息行为的引导和监管对于维护国家安全稳定有着重要意义。因此，对重大突发事件中网络社会群体信息行为主题图谱及演化的相关研究，是当前学术界和应急管理相关部门应关注的重点问题。

目前，国内外学者对重大突发事件下的社交媒体信息传播开展了相关研究。

* 本文系国家社会科学基金重大项目“大数据驱动的社交网络舆情主题图谱构建及调控策略研究”（项目编号:18ZDA310）研究成果之一。

作者简介：王晰巍(ORCID:0000-0002-5850-0126),教授，博士生导师，吉林大学大数据管理研究中心主任，吉林大学网络空间治理研究中心教授，E-mail:wxw_mail@163.com；李聪聪(ORCID:0009-0004-1408-8098)，吉林大学商学与管理学院硕士研究生，E-mail：lcc661089@163.com；孙哲(ORCID:0009-0009-0319-3665)，吉林大学商学与管理学院本科生，E-mail：jlusunzhe@163.com；李玥琪(ORCID:0000-0002-5930-7213)，吉林大学商学与管理学院博士研究生，E-mail：yueqili0407@163.com。

国外有学者对社交媒体平台数据的文本复杂性进行实证分析，据此研究社交媒体在应急响应领域所发挥的作用[2]；结合公众、应急组织和学术机构在重大突发事件中对社交媒体工具的使用情况，分析重大突发事件中社交媒体的作用机制[3]；通过机器学习技术，研究社交媒体平台在突发事件中加强或削弱风险认知方面的作用[4]；还有学者针对重大突发事件中政府部门发布的相关政策，通过数据分析技术，研究政府通过社交媒体进行信息发布对公众负面情绪传播及演化的影响[5]等。国内有学者总结了重大突发事件下社交媒体信息发布模式，并据此构建政府部门和医疗专家不同渠道信息发布的演化博弈模型[6]；还有研究应用质性分析方法从多维度对重大突发事件下的社交媒体信息过载的成因和影响展开[7]；分析重大突发事件下社交媒体多模态信息的特征，并归纳总结出多模态信息分析的关键方法和技术[8]；通过采用扎根方法和主题挖掘，研究基于数据分析的全媒体融合对重大突发事件下应急管理不同发展阶段的应用情况[9]。从国内外现有研究成果的现状来看，国外研究多关注重大突发事件中社交媒体的作用机制以及完善现有的应急管理系统等，国内相关研究多对重大突发事件下社交媒体信息传播模型进行构建和信息传播特征进行分析。从现有研究成果来看，国内外针对重大突发事件下社会群体信息行为演进相关研究的成果相对较少，针对社会群体信息行为舆情主题及演化特征的研究成果更少。

本研究试图解决以下两个研究问题：如何构建重大突发事件中社会群体信息行为舆情主题图谱，并对社会群体信息行为进行分析？如何针对重大突发事件中社会群体的信息行为进行网络舆情主题识别及舆情引导？本文研究在理论层面上，创新性地构建重大突发事件中社会群体信息行为舆情主题分析过程模型，明确主题聚类图谱、主题热度演化图谱和主题路径演化图谱的构建，为网络舆情的主题分析提供了新的理论框架和分析方法；在实践层面上，有利于营造健康良好的网络生态，为重大突发事件中社会群体信息行为舆情引导和舆情治理提供参考。

2 相关概念及文献综述

2.1 重大突发事件

重大突发事件是指突然发生，造成或者可能造成严重社会危害，需要采取应急处置措施并予以应对的自然灾害、事故灾难、公共卫生事件和社会安全事件[10]。重大突发事件主要有突发性、特定性、复杂性和危害性四个特征。重大突发事件的发生会影响社会稳定，造成人民恐慌。当重大突发事件发生后，事物原有的发展格局突然被打乱，可能会出现超出正常社会秩序和人们的心理惯性运行。因此，突发性事件出现后人们往往措手不及难以应对，整个工作和生活秩序混乱，不仅对世界各国的经济政治造成严重影响，甚至会给世界各国人民都带来深重灾难，严重影响各国的经济政治秩序和人民的正常生活。二十大报告中习近平总书记立足国情，多次强调“完善公共安全体系，提高防灾减灾救灾和急难险重突发公共事件处置保障能力”，完善国家应急管理体系建设，为重大突发事件的应急管理和舆情疏导工作指明方向[11]。

2.2 社会群体信息行为

信息行为是指个人在满足自己的信息需求过程中所发生的行为活动，主要是指以某种方式搜寻、使用和传递信息的相关活动[12]。国内外对重大突发事件下群

体信息行为的研究主要集中于构建群体信息行为模型，发现群体信息行为动机、过程机理及群体信息行为影响因素。初期研究多是采纳不同领域模型如钩状模型[13]和 TTM 模型[14]，从而对群体信息行为的影响因素进行分析，并基于影响因素模型提出对群体信息行为过程的服务优化。随后出现了通过采纳不同的需求理论对社交媒体中用户的信息行为机理、过程及行为动机进行解读的研究[15]；还有学者通过定性分析的方法确定用户信息行为过程中不同影响因素间的内在联系及作用路径[16]。随着互联网信息技术的快速发展，网络用户群体数量日趋庞大，有关社会群体信息行为的研究也逐渐延伸到互联网世界当中。本文研究的社会群体信息行为是指网络用户群体的信息行为，即网络用户群体在网络社交平台中进行信息获取、信息检索、信息共享、信息传播和信息利用等活动。这种信息行为具有开放性、无组织性、虚拟性和个性化等特点，大多数网络社交平台都允许网络用户在遵守网络道德规范和法律法规的前提下，在任何时间地点自由地进行任何信息活动。

2.3 主题聚类图谱

主题图谱是知识图谱在领域层面的应用及深化，它存在多种类别，包括用户交互行为图谱、语义图谱和事理图谱等[17]。其中，用户交互行为图谱是对用户实体及实体关系进行构建，旨在发现用户之间交互关系种类及网络中的重要节点等信息的关系。在行为图谱中，实体是指参与话题讨论的用户，每个实体具有一定的属性值，包括网络用户基本信息，如年龄、地域和认证信息等；实体与实体之间的关系表现为转发、评论、提及和点赞等关系。本文研究的主题聚类图谱，是在用户交互行为图谱的基础上，将海量的微博用户发布的评论转发文本信息通过主题相似度进行划分，不仅可以将社交网络中的用户通过主题分类成不同的聚类群体，还可以将相同主题下的网络用户聚集在一起，表现出不同网络用户群体对某一话题事件的转发、评论和提及行为[18]。主题聚类图谱采用图的建模方式，实体为参与话题讨论的网络用户，实体关系为不同用户之间的转发、评论和提及等。网络社交媒体中的主题聚类图谱来源于网络用户的评论转发文本信息，主题聚类图谱的构建有利于获取社交网络中的有效信息，如网络用户讨论的热点主题以及用户意见领袖识别等。

3 研究设计

3.1 基于 LDA 的社会群体信息行为舆情主题聚类演化图谱

通过对社会群体信息行为舆情的主题聚类，可以从更加全面和准确的信息视角了解在重大突发事件下舆情事件的发生、演化和影响。同时，通过对舆情主题聚类图谱的分析，可以了解用户群体对舆情事件的关注点，从而准确把控舆情发展方向，更好地引导网络舆情，避免舆情事件产生的负面影响。

本文通过 LDA 模型进行主题聚类。LDA 主题模型是一种可对大量文本信息特征进行主题建模的三层级贝叶斯概率图模型[19]，其组成结构包含文档、主题和词三种粒度。LDA 主题模型主要用于挖掘文本的潜在主题，并进行主题划分，集中在每篇文本的主题以概率分布的形式给出，并根据主题进行聚类。LDA 主题模型不考虑文档中词语的顺序，通常使用词袋特征（Bag-of-Word Feature）来代表文档，构成“文档-主题分布”和“主题-词分布”。

本文采用困惑度（perplexity）评价指标来确定文档中最优的主题个数。困惑度是一种信息理论的测量方法，在自然语言处理中困惑度常作为衡量一个概率分布或概率模型预测样本优劣程度的指标，可以通过困惑度数值变化来调节语言模型主题个数[20]，计算公式如下：

$$\mathrm{perplexity(D)} = \exp\left\{\frac{-\sum_{d=1}^{M}\log\left(P(W_d)\right)}{\sum_{d=1}^{M}N_d}\right\} \qquad 公式(1)$$

其中，D 表示所有文档的集合，M 表示文档的数量，Wd 表示文档 d 中的词，Nd 表示文档 d 中的词数，P(Wd)表示文档集合中词出现的概率。困惑度数值一般随着潜在主题数量的增加呈现递减的规律，困惑度数值越小表示该主题模型的生成能力越强，模型越好。因此，本文选择困惑度相对小且主题数量相对较少的主题数值，作为 LDA 模型训练的最优模型参数[21]。

3.2 基于时序性的社会群体信息行为舆情主题热度演化图谱

舆情主题热度演化是指舆情主题被公众关注的热点程度，随着时间的推移呈现出爆发或者消亡等变化趋势。对舆情事件主题热度演化的分析有助于提高对舆情事件生命周期判断的准确性，同时发掘出群体在舆情事件发展过程中受关注程度。通过 LDA 主题模型完成用户评论转发文本内容的分类主题挖掘后，截取舆情事件发生后的一段时间作为分析时间，对该段时间内的不同主题的文本数量进行统计[22]，从而作为舆情主题热度演化分析的衡量依据。

关于重大突发事件中社会群体信息行为的主题热度演化分析，本文基于主题热度和时序性，对不同舆情主题下的时间切片内出现的主题频次进行求和，以此来反映该主题在不同时段内的舆情主题讨论热度，即以发文时间为横轴，不同主题发文热度为纵轴，构建主题热度的演化图谱。通过针对主题热度演化趋势的分析，进行重大突发事件中社会群体信息行为主题在整个舆情事件发展过程中的演化特征分析。

3.3 基于相似度的社会群体信息行为舆情主题路径演化图谱

通过分析社交网络中舆情主题路径演化，可以及时发现和预警潜在的舆情事件，从而更好地应对和处理可能出现的负面影响。同时，研究社交网络中舆情主题路径演化，可以帮助国家政府更好地了解用户的信息需求和传播路径，从而优化信息传播策略，提高信息传播效果。

社交网络具有以下特点：①在社交网络中，意见领袖节点往往代表其所在社会群体的主题倾向；②社交网络中的边权重可以看作是网络社群在不同节点间传播的信息损耗，即节点的主题相似度越高，它们之间的信息传播就更为容易，信息损耗也越小[23]。因此，为降低后续计算的复杂度，保障研究的科学性，本文首先挖掘不同群体意见领袖的代表其所在的群体，再计算出群体意见领袖之间的相似度代表主题之间的语义距离，并将语义距离作为主题的边权重，最后计算遍历各舆情主题的最短路径，构建社会群体信息行为舆情主题路径演化图谱。

PageRank 算法是由谷歌公司创始人拉里·佩奇和谢尔盖·布林于 1996 年推出的一种网页分析算法[24]。原本的 PageRank 算法模型主要用于确定网页搜索排名情况，根据网页的入度和出度来计算网页的权重，权重越高的网页排名越靠前。而近来，大量关于 PageRank 算法的研究都致力于使用 PageRank 算法度量社交网

络中意见领袖的影响力[25]。由于 PageRank 算法不仅考虑了节点的数量和质量，也考虑了不同节点之间的关联，是一种比较综合的排名算法，因此本文选用 PageRank 值来挖掘重大突发事件中不同社会群体的意见领袖。

KL 散度也称相对熵，是两个概率分布间差异的非对称性度量，计算见公式（2）。JS 散度是在 KL 散度的基础上演化而来的，是用于描述两个概率分布的对称性相似程度[26]，计算见公式（3）。

$$KL(P||Q)=\sum P(x)\log\frac{P(x)}{Q(x)} \qquad \text{公式(2)}$$

$$JS(P||Q)=\frac{1}{2}KL\left(P(x)||\frac{P(x)+Q(x)}{2}\right)+\frac{1}{2}KL\left(Q(x)||\frac{P(x)+Q(x)}{2}\right) \qquad \text{公式(3)}$$

从公式（3）可以看出等式是对称成立的，也就是说 $JS(P||Q)=JS(Q||P)$。在本文中，P(x)和 Q(x)表示不同意见领袖的概率分布，即 LDA 主题模型计算出的“文档-主题”分布，JS 散度的值域范围是[0,1]，P(x)和 Q(x)的相似度越高，JS 散度值越接近于 0[27]。因此，本文选择 JS 散度来度量不同意见领袖之间的相似度。

深度优先搜索（DFS）算法是一种基于栈或递归实现的搜索算法，其基本思想是从起点出发，依次访问相邻的节点，直到找到目标节点或无法继续访问为止。在访问一个节点时，如果该节点未被访问过，则标记为已访问，并将其加入栈中，然后继续访问与该节点相邻的节点。如果所有相邻的节点都已被访问过或者没有相邻节点，则从栈中弹出上一个节点，回溯到上一个节点，继续访问其它未访问过的相邻节点[28]。DFS 算法可以用来解决许多图论问题，如连通性、最短路径、最小生成树等问题。因此，本文通过 DFS 算法来计算遍历各舆情主题的最短路径。

3.4 重大突发事件中社会群体信息行为舆情主题图谱构建过程模型

基于上述分析，本文构建了重大突发事件中社会群体信息行为舆情主题图谱及演化特征分析过程模型，如图 1 所示。过程模型分为 4 个阶段：数据处理、主题挖掘、主题图谱构建和舆情主题分析。（1）数据处理。通过软件和编码对重大突发事件下网络舆情事件数据进行爬取，爬取的数据包括用户名、用户评论转发文本、评论转发关系和发文时间，随后对爬取到的数据进行预处理，即清洗、分词、筛选和去重等操作，获得预处理后的数据。（2）主题挖掘。通过困惑度评价指标确定最优主题个数，再利用 LDA 主题模型对预处理后的数据进行主题挖掘，随后进行主题时序统计和主题频次统计。通过 PageRank 值识别不同主题意见领袖，结合 LDA 主题模型得出的“文档-主题分布”，计算出不同群体意见领袖之间的 JS 散度值作为主题之间相似度，通过 DFS 算法来计算遍历各舆情主题的最短路径。（3）主题图谱构建。通过编码和可视化软件构建社会群体信息行为舆情主题聚类图谱、主题热度演化图谱和主题路径演化图谱。（4）舆情主题分析。分别对主题聚类图谱、主题热度演化图谱和主题路径演化图谱进行主题特征分析、主题热度分析和主题最优传播路径分析。

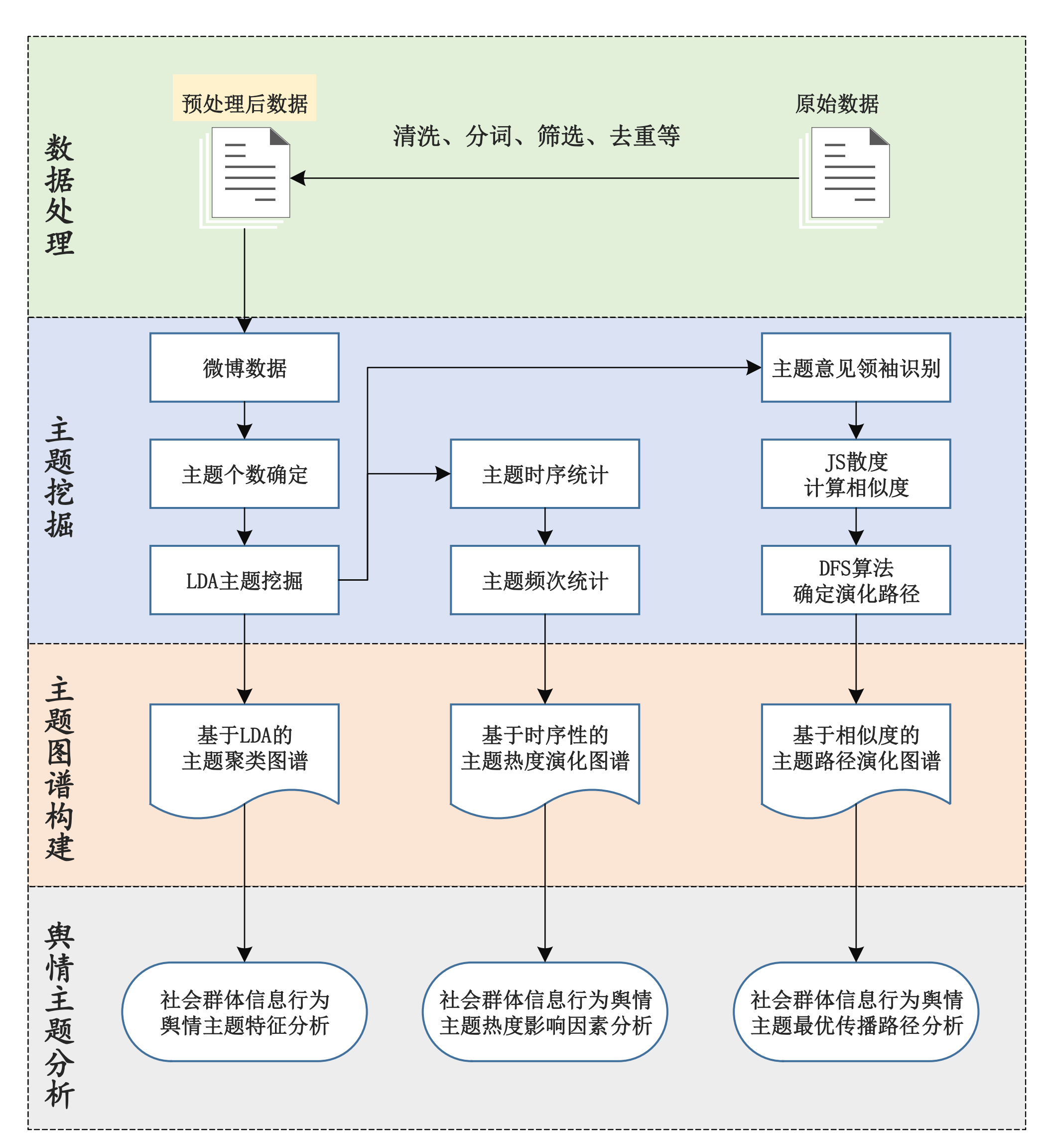


图 1 重大突发事件中社会群体信息行为舆情主题图谱构建过程模型

4 实证分析

4.1 数据处理

据新浪微博 2022 年 3 月公布的《新浪微博 2021 年全年财报》显示，截至 2021 年末微博月活跃用户达到 5.73 亿，同比增长 10%[29]。微博平台作为目前国内主流社交媒体平台之一，具有用户门槛低、信息时效性强和传播范围广泛等特点。2022 年 3 月 21 日，中国东方航空集团有限公司 MU5735 航班在广西壮族自治区梧州市附近山林坠毁，发生重大空难事故。此次重大突发事件，给国家和人民都带来了难以估量的危害，不仅造成严重财产损失和人员伤亡，也使得我国民航业再次遭受重创。随着“3.21”东航客机事故的持续发展，其造成的严重人员伤亡、灾后救援和事故原因等相关话题迅速引起网络社交平台中大量官方媒体和普通用户的关注。

因此，本文选择新浪微博作为研究平台，选取新浪微博的“3.21”东航客机事故这一具有一定代表性重大突发公共安全事件作为舆情话题，采用八爪鱼工具对舆情事件下的数据进行爬取，收集的数据信息主要包括用户名、用户 ID、发

文时间、评论转发文本内容等。对收集的数据进行数据处理。首先使用 Microsoft Excel 软件筛查并剔除掉无关数据，去除表情及链接等。其次，调用 Python 3.10.6 中的 jieba 库对文本进行分词操作，然后筛选出文本中的名词、动名词、专有名词、形容词作为文本关键词，再人工删除无关词、虚词和停用词，如“公司”、“代表”等。再次，对同义词进行合并，如将“家人”、“亲人”都合并为“家属”，并创建常用词典合成常用词，如 jieba 库可能将“央视”和“新闻”分成了两个词，需要在常用词典中将其合并成“央视新闻”这一词。最后，得到微博评论转发数据共 28326 条。

为更直观地呈现出预处理后的数据结果，发现关键词之间的潜在关系和微博用户关注的热点话题，深入了解用户群体对舆情事件的态度和情感，本文在 Windows10 系统下，通过 Python 3.10.6 编码和 Microsoft Excel 软件，对数据处理后的关键词进行筛选，筛选出词频最高的 200 个关键词，构建社会群体信息行为舆情关键词共现矩阵，通过共现矩阵构建共现网络。

Gephi 是一款开源的网络分析和可视化工具，它可以通过导入节点及边表格或导入矩阵帮助用户对复杂的网络数据进行分析和可视化。本文选择目前的最高版本 Gephi 0.9.7 构建“3.21”东航客机事故关键词共现网络并进行后续分析。通过 Gephi 可视化软件对“3.21”东航客机事故社会群体信息行为舆情关键词共现网络进行可视化呈现，如图 2 所示。以关键词为共现网络中的节点，节点大小代表关键词的度中心性，共现关系为网络中的边，边的粗细代表关键词之间的共现次数，共获得 200 个节点和 5070 条边。

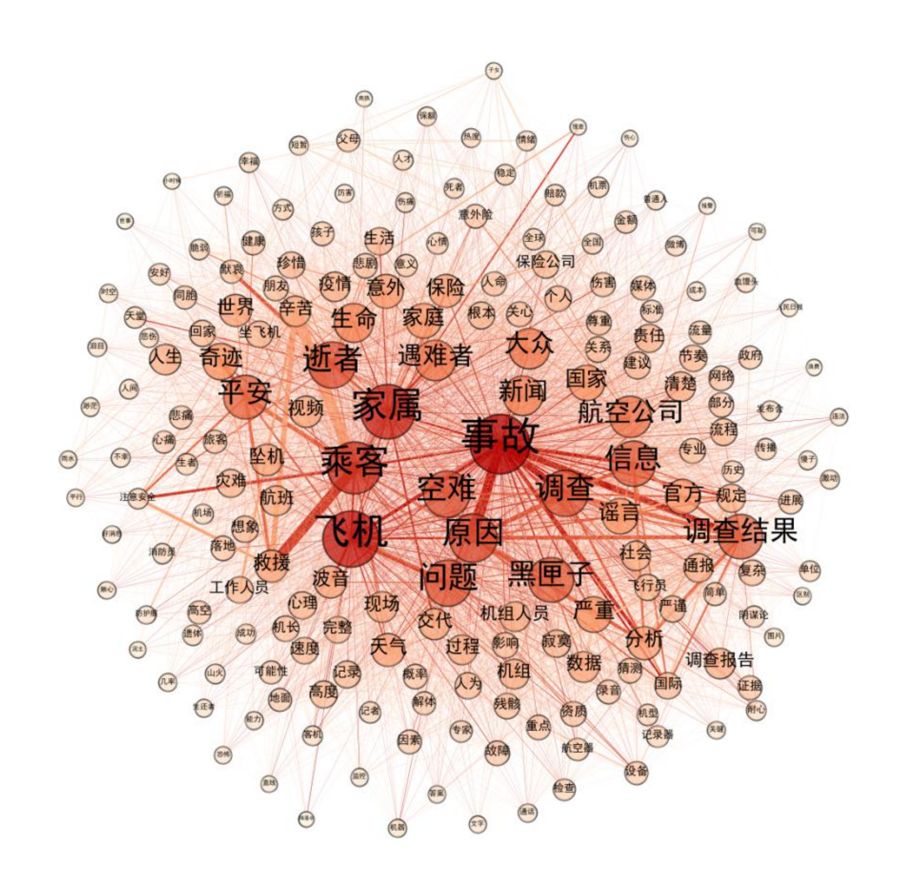


图 2 “3.21”东航客机事故关键词共现网络

社会群体信息行为舆情关键词共现网络，可以展示各群体信息行为舆情关键词重要性程度，也可以展示各关键词之间的关联关系，还可以根据边的权重展示关键词之间的关联关系程度。根据以往研究中对社会网络的分析方法，本文从网络密度和网络中心性两个角度对“3.21”东航客机事故关键词共现网络进行分析。一方面，选择网络密度为分析指标，对关键词共现网络进行整体分析；另一方面，选择网络中心性为分析指标，对网络结构中度中心性为前 10 的节点为代表性节点进行重点分析，此类节点是影响舆情事件发展的关键节点及关键要素。

通过 Gephi 测算“3.21”东航客机事故关键词共现网络密度，可知“3.21”东航客机事故关键词共现网络密度为 0.255，几乎接近 0。一般来说网络密度越大，各节点的互动越频繁。这意味着“3.21”东航客机事故在新浪微博上传播时社交媒体用户节点之间交流相对较少，致使用户在接收信息时需要对舆情事件按时间顺序进行梳理，所以花费了较多时间来全面掌握当前这一重大突发安全事件的发展态势。同时，本文通过 Gephi 分别测算“3.21”东航客机事故关键词共现网络中点度中心性、中介中心性和接近中心性前 10 的节点，进行网络中心性指标的详细分析，如表 1 所示。

表 1 “3.21”东航客机事故关键词共现网络中心性指标详细信息

序号	Label	degree	Label	Betweenness Centrality	Label	Closeness Centrality
1	事故	184	事故	1201.43	事故	0.930
2	飞机	172	飞机	990.32	飞机	0.881
3	家属	163	家属	881.77	家属	0.847
4	乘客	156	乘客	686.58	乘客	0.822
5	原因	140	逝者	522.84	原因	0.771
6	逝者	135	平安	459.54	逝者	0.757
7	空难	134	信息	449.87	空难	0.754
8	问题	133	原因	445.33	问题	0.751
9	调查	133	问题	442.92	调查	0.751
10	信息	128	空难	424.73	信息	0.737

4.2 数据结果

研究结果表明，“3.21”东航客机事故舆情事件的关键词共现网络通过对网络密度及代表性节点中心性的分析，发现重大突发事件下社会群体信息行为具有以下几个特征：

重大突发事件下社会群体信息行为舆情传播主体具有很强的交互性和关联性。通过对关键词共现网络的分析，发现不同关键词节点之间边的粗细有所不同，粗细程度代表关键词节点之间的共现次数，即节点之间的边越粗，节点之间的共现次数越多，节点关系越密切。例如，“事故”“原因”“调查”“调查结果”“信息”等关键词节点之间的边较粗，说明这些节点之间的联系较为密切，体现了重大突发事件中社会群体十分关注事故调查结果，渴求知晓事发原因；关键词节点“乘客”与关键词节点“救援”、“平安”之间的边较粗，体现了重大突发事件中社会群体对受灾乘客的强烈忧心和殷切期盼。

重大突发事件下社会群体信息行为舆情传播网络具有一定的稀松性和异质性。从网络密度分析来看，“3.21”东航客机事故关键词共现网络的整体密度偏低为 0.255。由此可见，“3.21”东航客机事故在新浪微博平台传播时，各舆情传播社交网络的用户节点之间关注的主题关键词间互动联络都极为稀松，微博用户信息交互行为匮乏、关系不够紧密，各节点存在严重的异质性，难以形成一致的表达，也降低了信息的传播速度。

重大突发事件下社会群体信息行为舆情传播的关键节点具有一定的集中性。从网络中心性分析来看，在“3.21”东航客机事故中，点度中心性前 10 的关键词为“事故”“飞机”“家属”“乘客”“原因”“逝者”“空难”“问题”“调查”“信息”，说明这些关键词在信息传播网络中具有极高的地位；中介中心性前 10 的关键词为“事故”“飞机”“家属”“乘客”“逝者”“平安”“信息”“原因”“问题”“空难”，说明这些社交网络的节点发表的舆情主题关键词作传播能力较强，更能引导舆情的传播；接近中心性前 10 的关键词为“事故”“飞机”“家属”“乘客”“原因”“逝者”“空难”“问题”“调查”“信息”，说明这些社交网络的用户关键词到达其他关键词节点较为简单。由此可见，关键词“事故”“飞机”“家属”“乘客”“逝者”“空难”等均靠前排列，这些关键词都是信息传播网络中的关键节点，与其他关键词节点联系密切，是重大突发事件中舆情主体传播的核心关键词。“事故”“空难”是此次重大突发事件的属性，“飞机”“乘客”都是此次事故发生的主体，“家属”“逝者”则是事故发生后社会群体的反应，表明社会群体关心事故遇难的“逝者”及其“家属”，“3.21”东航客机重大突发安全事故引发全国民众的热切关注。

4.3 基于 LDA 的主题聚类图谱

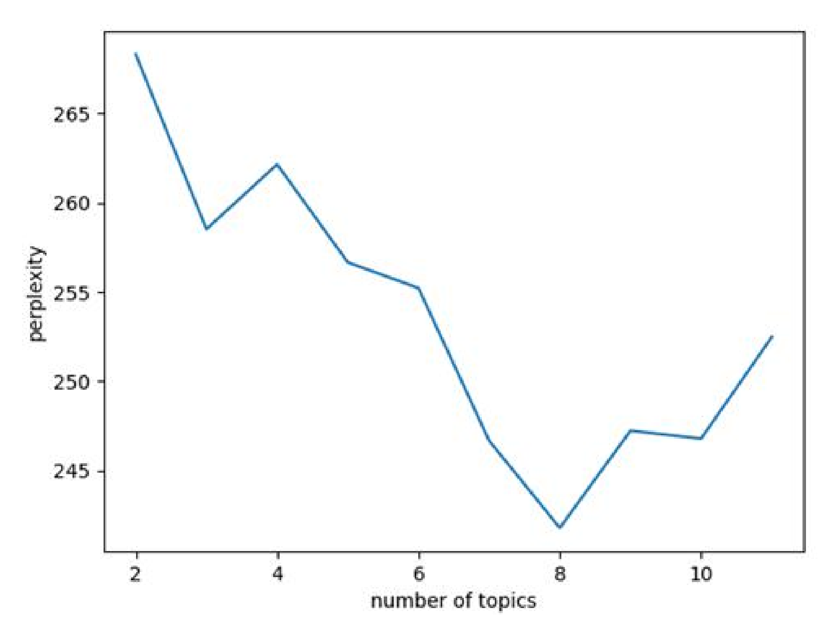


图 3 perplexity-topic 折线图

确定主题聚类个数。经过数据预处理后，本文选用基于 Python-sklearn 中的 LDA 主题模型，对预处理后的文本进行分类训练。主题个数太少可能不足以捕捉到数据的丰富语义结构，导致模型的表现较差。通常，至少需要两个主题才能进行对比和观察不同主题之间的关系[30]。而主题个数太多会导致模型过于复杂，

并且可能出现过拟合现象[31]，选择一个较小的主题个数范围可以帮助避免这种问题。因此，拟定选择区间[2,12]内的某一整数作为候选主题数。通过调用 LDA 主题模型类下的 Perplexity 方法，得出不同模型的困惑度数值，如图 3 所示。该图表明某一文档对于各个潜在主题的不确定程度。困惑度越低，文档归属于某一潜在主题的可能性就越高，即模型的聚类效果越好。图 3 的折线图显示，随着主题数的增加，总体上困惑度呈现先波动下降再上升的态势；困惑度的局部极小值点，出现在主题数为 8 的模型选择上。因此，本文舆情话题拟选取 8 个潜在主题数。

构建主题聚类图谱。本文的研究对象是新冠肺炎疫情舆情空间下的微博用户节点，任意两个节点间的转发评论关系，代表着两个节点在同一舆情空间下的话题倾向性趋同。因此，这种转发评论关系代表节点之间的相似性。本文以“3.21”东航客机事故舆情话题中微博用户为节点，转发评论为边，将最终获得的 28326 条微博转发评论数据导入 Gephi 0.9.7 可视化软件中，构建“3.21”东航客机事故微博用户主题聚类图谱，如图 4 所示。

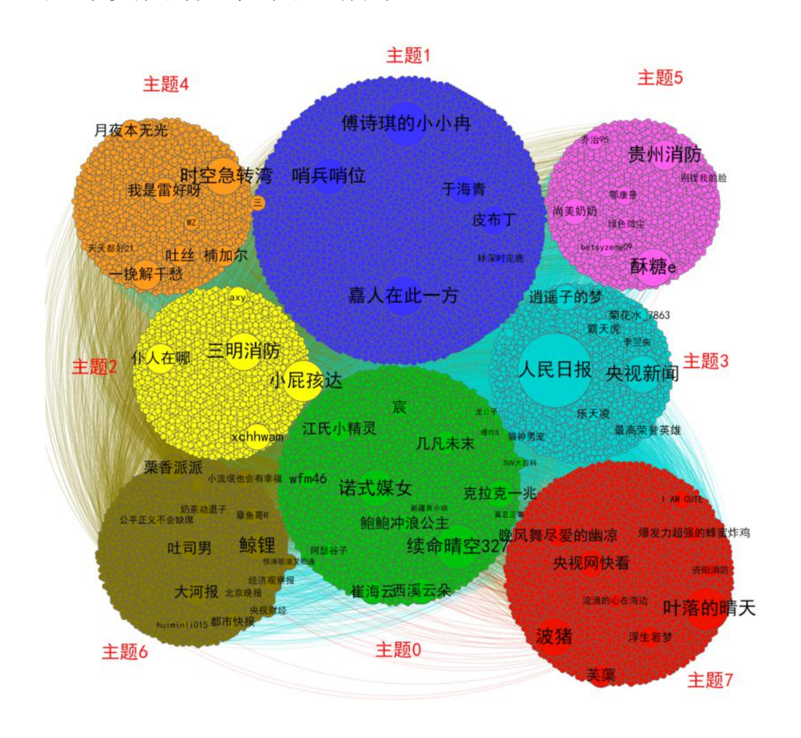


图 4 “3.21”东航客机事故微博用户主题聚类图谱

根据不同颜色，将“3.21”东航客机事故话题下的微博用户主题聚类划分为不同网络社群。其中，节点大小与节点的度中心性成正比。节点越大，其度中心性越大，说明该微博用户在重大突发事件下社会群体中的影响具有更高的地位和社群影响力。

在确定最优主题数后，将分词后的文本数据用于 LDA 主题模型进行训练，得到“主题-词”以及“文档-主题”两个概率分布。通过“主题-词”分布可确定各个主题包含的高频关键词，并以分类的主题个数确定微博用户群体。利用 LDA 主题模型训练得到的 8 个主题，且各个主题均选取词频最高的前 5 个词，并对主题进行人工归纳和命名，如表 2 所示。

表 2 主题高频关键词分布

主题 0：网络谣言识别概率	谣言 0.095	真相 0.058	不传谣 0.047	通报 0.035	官方 0.032
主题 1：祈祷奇迹降临概率	平安 0.370	奇迹 0.186	保佑 0.180	祈祷 0.042	坚强 0.020
主题 2：哀悼遇难者概率	逝者 0.091	难过 0.063	安息 0.035	痛心 0.019	揪心 0.019
主题 3：探究事故致因概率	调查结 0.113	调查 0.078	信息 0.069	原因 0.038	黑匣子 0.036
主题 4：意外险赔付概率	保险 0.052	事故 0.036	意外险 0.031	赔付 0.026	航空公司 0.022
主题 5：宽慰遇难者家属概率	家属 0.096	生命 0.053	好好活着 0.039	默哀 0.038	加油 0.025
主题 6：保障乘客安全概率	问题 0.111	飞机 0.106	波音 0.028	安全 0.024	乘客 0.016
主题 7：乘客救援概率	希望 0.216	乘客 0.071	信息 0.070	找到 0.045	救援 0.044

通过表 2 可以看出，各个主题的关键词都占有较大的概率值，这符合微博文本主题的特点，即微博用户在某一特定话题空间下的评论用词习惯趋于相同[32]。同时，除了极少数不同主题高频关键词相同，各个主题的高频关键词大体上各不相同，也说明了该模型能够较好地实现微博用户评论转发文本主题的划分。通过“主题概率分布”得出用户转发评论文本信息的主题划分，从而确定微博用户群体，并统计“3.21”东航客机事故微博用户群体主题概率分布，如图 5 所示。从图 5 中可以看出，主题出现频率从高到低依次为主题 1、主题 0、主题 7、主题 6、主题 4、主题 3、主题 2 和主题 5。其中，主题 1 占比最高达到 24.16%；主题 5 的占比最少，占比 8.41%左右。

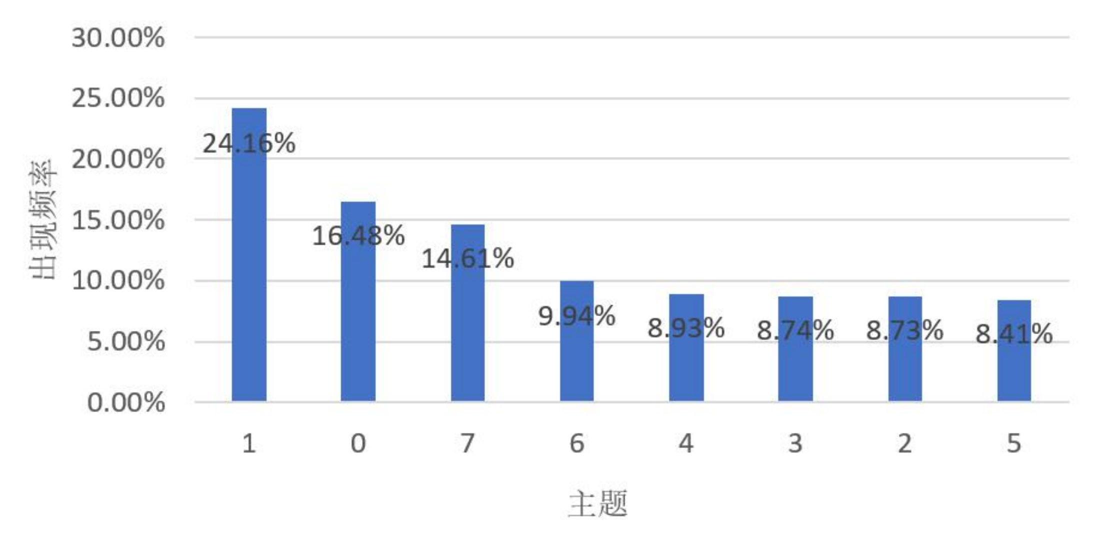


图 5 主题概率分布

4.4 基于时序性的主题热度演化图谱

基于 LDA 主题模型对重大突发事件中社会群体信息行为进行主题划分后，本文基于主题热度和时序性，对不同主题和不同时间切片内出现的频次进行求和，来反映不同舆情传播主题在不同时段内的讨论热度，以此构建重大突发事件下的社群主题热度演化图谱。考虑到舆情事件的生命周期，事件爆发较长时间之后产生的数据量一般较少，研究价值不大，故本研究仅以“3.21”东航客机事故发生后一个月左右作为分析时间，截取 2022 年 3 月 21 日到 2022 年 4 月 23 日期间的微博用户评论转发文本作为研究数据。本文选择主题河流图来展示不同时间不同主题的演化情况，通过调用 Python-pyecharts 工具包构建主题热度演化图谱，其中发文时间为横轴，舆情关注的主题热度为纵轴，不同颜色代表不同发文主题，不同主题河流的宽度代表对应时间点的主题热度，即主题河流所占纵轴比例越大其讨论热度越高。据此构建了“3.21”东航客机事故主题热度演化图谱，见图 6。

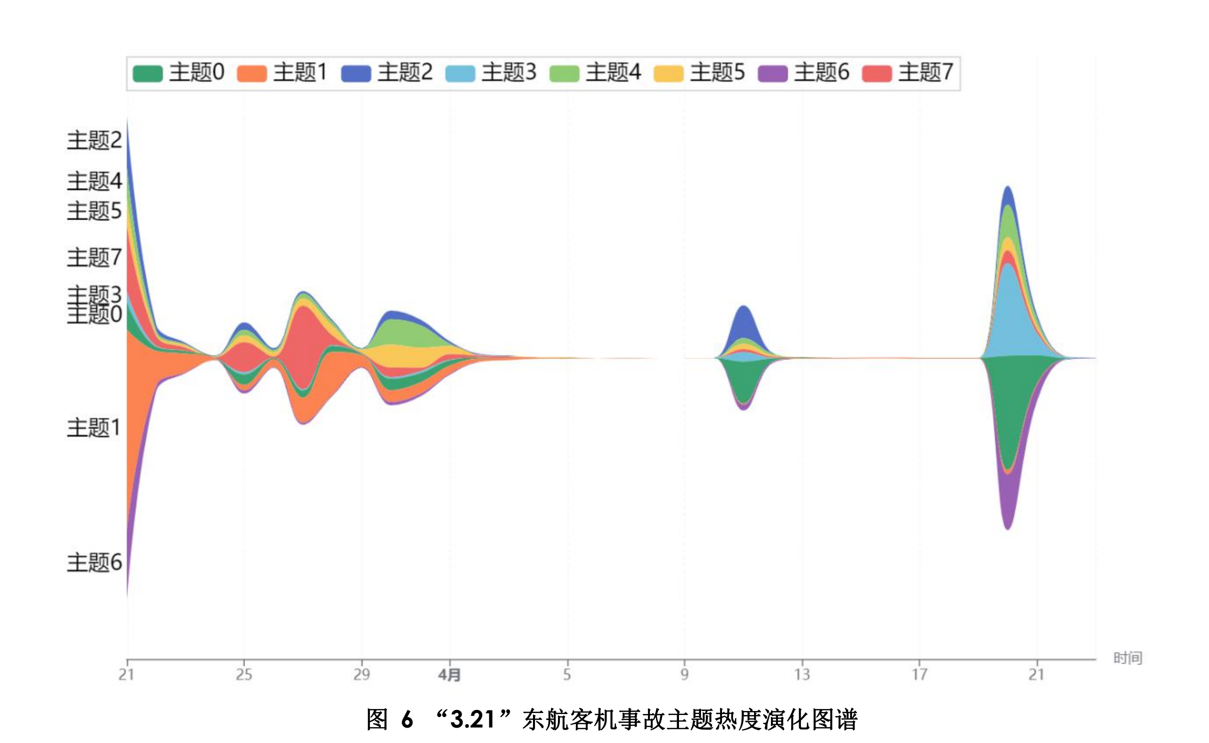


图 6 “3.21”东航客机事故主题热度演化图谱

4.5 基于相似度的主题路径演化图谱

识别不同主题群体的意见领袖，代表其所在主题的社交网络群体。PageRank 算法基于网络图模型，通过综合每个节点的点度中心性和网络传播特性来确定该节点的重要程度，即节点的 PageRank 值越高，表示该社交网络的节点对社群的影响力越大。本文在对主题进行划分之后，挖掘各主题群体中 PageRank 值最大的节点作为其所在主题群体的意见领袖，见表 3。

表 3 主题群体意见领袖

主题	意见领袖	PageRank 值
0	续命晴空 327	0.00878
1	嘉人在此一方	0.00526
2	三明消防	0.00384
3	人民日报	0.06784
4	时空急转湾	0.00415
5	酥糖 e	0.00325
6	鲸锂	0.00239
7	叶落的晴天	0.00527

在识别出各主题群体的意见领袖之后，结合 LDA 主题模型得出的“文档-主题分布”，计算出不同群体意见领袖之间的 JS 散度值作为主题之间相似度，代表不同主题之间的边权重，见表 4。

表 4 主题群体意见领袖相似度

主题	0	1	2	3	4	5	6	7
0	0	0.2587	0.4007	0.1360	0.1548	0.3417	0.0988	0.1091
1	0.2587	0	0.2574	0.3334	0.2702	0.3370	0.2394	0.2326
2	0.4007	0.2574	0	0.3556	0.3797	0.1564	0.2935	0.2255
3	0.1360	0.3334	0.3556	0	0.2342	0.3598	0.1725	0.2047
4	0.1548	0.2702	0.3797	0.2342	0	0.3930	0.1683	0.1459
5	0.3417	0.3370	0.1564	0.3598	0.3930	0	0.2713	0.1652
6	0.0988	0.2394	0.2935	0.1725	0.1683	0.2713	0	0.0405
7	0.1091	0.2326	0.2255	0.2047	0.1459	0.1652	0.0405	0

根据不同主题群体意见领袖的相似度，确定不同主题意见领袖之间主题演化路径上的边权重，以最大主题群体 1 作为主题路径演化的起点，最小主题群体 5 为主题路径演化的终点，通过 DFS 算法计算遍历各舆情主题的最短路径，得到主题演化路径为“主题 1->主题 4->主题 0->主题 3->主题 6->主题 7->主题 2->主题 5”。据此构建重大突发事件下社会群体信息行为舆情主题路径的演化图谱，见图 7。由图 7 可知，在选择主题群体 1 进行舆情信息的传播时，从最大主题群体 1 出发，最后到达最小主题群体 5 的主题传播路径，所带来的信息损耗最小。通过确定不同主题舆情传播的最短路径，可以有效减少舆情传播过程中的信息失真，同时可以更加精确地掌握舆情信息的传播情况，及时发现和纠正谣言等负面信息，避免其不断发酵。因此，识别舆情主题意见领袖，构建舆情事件主题路径演化图谱，可以减少舆情在传播过程中的信息损耗，提高舆情管理的效率和准确性。

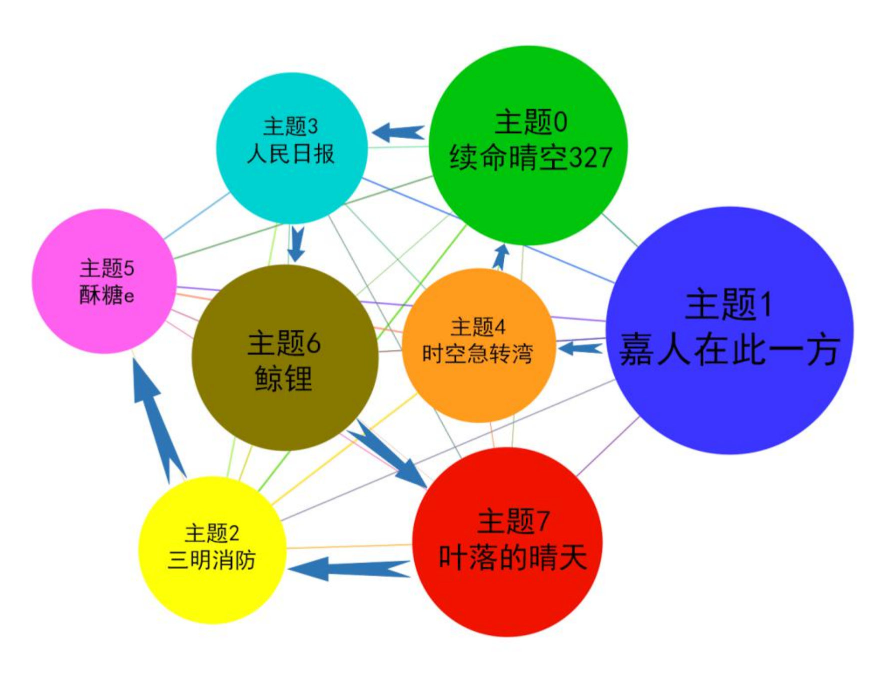


图 7 “3.21”东航客机事故主题路径演化图谱

5 讨论分析

5.1 重大突发事件中社会群体信息行为舆情主题特征分析

根据用户群体 LDA 主题聚类图谱、主题高频关键词分布和主题概率分布，发现“3.21”东航客机事故网络舆情发展周期内出现的社会群体主题。按照主题频率由高到低进行分析，可以发现：主题 1 的用户群体仍然抱有一丝希望，希望乘客能够生还。他们祈求奇迹出现，并期待调查结果能够给出好消息。主题 7 的用户群体在事故发生后第一时间内，就呼吁政府部门加紧救援，希望专家能够尽快确定飞机失事所在区域。他们对乘客的安全非常关注，希望救援行动能够尽快展开。主题 2 的用户群体感到十分痛心，对失事的航班表示哀悼，并祈愿逝者安息、生者坚强。他们希望这样的事故不再发生，同时也要关注乘客家属的情况，尽力安抚他们的情绪。主题 5 的用户群体关注失事飞机乘客的家属情况，希望他们能够节哀顺变，珍惜生命，保重身体健康。主题 6 的用户群体对波音系列客机的安全性持怀疑态度，认为乘客的人身安全未必能够得到保障。他们希望相关部门能够对波音系列客机进行更加严格的检查，保障乘客的人身安全。主题 3 的用户群体主要关注此次事故的调查结果，探寻酿成悲剧的主要原因。他们希望相关部门能够尽快公布调查结果，为事故的原因找到合理的解释。主题 4 的用户群体主要关注事故后意外保险等相关问题，以及航空公司的赔付情况。他们认为，保险公司和航空公司应该承担相应责任，为乘客及其家属提供必要的帮助和赔偿。主题 0 的用户群体面对繁杂的信息，保持理智，甄别网络谣言。他们耐心等待官方发布的调查结果，以便了解事故的真相和原因。

本文所构建的重大突发事件社会群体信息行为舆情主题特征分析，不仅能够直观呈现重大突发事件下用户之间关注的主题交互状况，同时通过 LDA 方法可以有效划分舆情主题并确保重大突发事件中舆情主题特征分析的有效性。以往研究中大多通过知识图谱[33]进行网络舆情的分析，也有一些学者通过事理图谱[34]对网络舆情进行研究。本文通过 LDA 主题聚类图谱的构建，发掘重大突发事件中社会

群体信息行为的舆情主题特征。一方面，可以通过对重大突发事件下舆情主题特征的分析，帮助舆情监管部门准确有效地识别微博用户群体话题，以更快地了解重大突发事件下社会群体关注的重要主题方向和舆情争议的焦点，获取公众声音和意见，以更好地进行舆情引导和情感疏导。另一方面，可以实时监测和预警重大突发事件的舆情动态，通过对舆情数据的分析和可视化，及时捕捉和分析舆情走向和进行舆情风险预警[35]，并结合不同主题的网络群体特征做到有针对性的舆情监管和引导，从而更好地实现对重大突发事件下舆情的引导和社会的稳定。

5.2 重大突发事件中社会群体信息行为舆情主题热度影响因素

根据“3.21”东航客机事故主题热度演化图谱，该事故从 2022 年 3 月 21 日东航航班机在广西梧州山林坠毁开始，是整个话题空间的入口。2022 年 3 月 21 日到 2022 年 4 月 23 日期间 8 个主题均出现了 5 个峰值，分别为 2022 年 3 月 25 日、2022 年 3 月 27 日、2022 年 3 月 30 日、2022 年 4 月 11 日和 2022 年 4 月 20 日。从单个峰值来看，乘客“救援主题”分别占据了 2022 年 3 月 25 日和 2022 年 3 月 27 日这两个峰值的最大讨论热度，“意外险赔付主题”占据了 2022 年 3 月 30 日这一峰值的最大讨论热度，“网络谣言识别主题”占据了 2022 年 4 月 11 日这一峰值的最大讨论热度，探究事故致因主题占据了 2022 年 4 月 20 日这一峰值的最大讨论热度。通过相关事件梳理，发现以下事件是导致这 5 个峰值产生的原因：2022 年 3 月 25 日东航客机坠毁事故救援进入第 5 天，在 25 日召开的“3·21”东航 MU5735 航空器飞行事故国家应急处置指挥部第五场新闻发布会上，公众关心的诸多焦点问题得到回应，诸如黑匣子破译和遇难者身份鉴定进展；2022 年 3 月 27 日，“3·21”东航 MU5735 航空器飞行事故遇难者集体哀悼活动在事发地搜救现场举行，对遇难者表示哀悼；2022 年 3 月 30 日，据中国银保监会消息称“东航遇难者家属已获 11 家保险公司赔付 1485 万元”；2022 年 4 月 11 日，官方辟谣东航飞机失事与副驾有关，并表示 MU5735 事故原因还在调查中；2022 年 4 月 20 日，官方公布针对东航 MU5735 航班失事的初步报告。

通过分析“3.21”东航客机事故主题热度的演化，本文进一步验证了影响网络舆情主题讨论热度的演化因素。以往研究对网络舆情主题热度影响因素的研究，多集中于认为标志性事件的出现是影响网络舆情主题热度演化的主要因素之一[36]，舆情事件本身的特性也对网络舆情主题热度产生较大影响[37]。本文在前序研究的基础上，进一步验证了官方相关信息公开时间和信息公开的范围、数量对网络舆情主体热度具有一定的影响。为加强真实信息的传播效率，官方机构应尽早进入网络信息公开系统，并向公众公开相关信息，安抚公众情绪。在此基础上，相关组织也需要建立谣言监测和预警机制。通过关注信息公开的范围及数量，确定官方信息渠道，如政府部门或权威机构发布及时通告和举行新闻发布会，确保重大突发事件信息传播的准确性和可信度。另一方面，确定影响力较高的谣言传播者并对其进行引导和管制，改变信息的流向和流速，有效避免负面信息的发酵，从而控制重大突发事件下社交媒体的网络谣言的传播速度和负面舆论倾向。

5.3 重大突发事件中社会群体信息行为舆情主题最优传播路径分析

本文根据数据分析结果与主题路径演化分析发现，在话题“3.21 东航客机事故”中，意见领袖“嘉人在此一方”→“酥糖 e” →“嘉人在此一方” →“时空急转湾” →“续命晴空 327” →“人民日报”→“鲸锂”→“叶落的晴天”→“三明消防”→“酥糖 e”的舆情传播路径最短，即舆情主题 1 到主题 5 的最

优传播路径为主题 1->主题 4->主题 0->主题 3->主题 6->主题 7->主题 2->主题 5。舆情主题通过该路径传播时造成的信息损耗最小、信息传播效率最高。在“3.21 东航客机事故”的重大突发事件的舆情话题中，可以通过确定不同主题群体间的最优演化路径，同时利用社交网络的连通性，进行更精准高效的主题群体意见领袖话题的推送，从而实现科学舆情监测和舆论引导。

重大突发事件舆情传播过程中，对于意见领袖的引导和管理发挥着至关重要的作用，他们的言论往往会引导舆情的走向。目前，多项研究考虑到意见领袖对于网络舆情管控的影响[38]，并有学者讨论意见领袖是信息传播效率内在驱动之一[39]。为了提高舆情管控的效果，需要采取“引导”与“控制”相结合的方式，让意见领袖发挥更大的作用。意见领袖的权威性和可信度是影响普通用户转发行为的关键因素[40]，官方机构可以选取影响力及权威度较高的意见领袖进行信息公开，以提升单个节点信息传播效率。此外，考虑到推送话题时，对所有用户进行无差别的推送容易造成推送阻塞或推送风暴，降低用户对推送内容的信任度，本文提出在舆情监管中，通过确定不同群体间主题传播的最优路径，利用社交网络的连通性，实现群体意见领袖的高效话题推送，以降低信息传播过程中的失真，更好地引导重大突发事件中网络舆情的走向。因此，构建重大突发事件中社会群体信息行为舆情主题路径演化图谱，识别舆情主题意见领袖，确定主题传播的最优路径，可以有效减少舆情传播过程中的信息损耗和失真，提高舆情管控的效率和准确性。

6 研究结论

本文在理论层面上，构建重大突发事件中社会群体信息行为舆情主题分析过程模型，明确基于 LDA 的主题聚类图谱、基于时序性的主题热度演化图谱和基于相似度的主题路径演化图谱的构建方法。研究结果表明，利用本文构建的主题聚类图谱、主题热度演化图谱和主题路径演化图谱，可以发现重大突发事件中社会群体信息行为的舆情主题特征、舆情主题热度影响因素以及舆情主题最优传播路径。本文的研究，为重大突发事件中社会群体信息行为舆情分析提供了新的理论框架和分析方法。

在实践层面上，以“3.21”东航客机事故舆情话题为例，通过重大突发事件中社会群体信息行为主题特征分析，帮助舆情监管部门准确有效地识别微博用户群体敏感舆情话题和敏感群体，从而有效确定在重大突发事件舆情传播中的重点舆情监管节点；通过对重大突发事件中社会群体信息行为舆情主题热度影响因素分析，可以识别敏感舆情主题信息的流向和流速，有效避免负面舆情的发酵，从而控制社交网络中舆情的信息传播速度、传播量和舆论倾向；通过对重大突发事件中社会群体信息行为舆情主题最优传播路径的分析，可以有效减少舆情传播过程中的信息传播失真，避免错误信息和网络谣言的传播，通过算法的优化推荐提高舆情管理的效率和舆情的有效引导。本文在实践层面可为舆情监管部门提供更好的社会群体舆情主题监管方法，可更好地推动重大突发事件下负面网络舆情的疏导，营造更好的网络生态环境。

本文在研究中存在一定的局限性。选取的微博平台虽然具有一定的代表性，但是不能涵盖其他舆情平台“3.21”东航客机事故话题内容。同时，仅结合“3.21”东航客机事故这一话题进行分析研究。后续将进一步补充更多代表性的舆情平台，

以及针对更多典型舆情话题进行重大突发事件中社会群体信息行为舆情主题图谱及演化进行研究。

参考文献

[1] 二十大报告全文来了！[EB/OL].[2022-10-25]. http://www.chinajsb.cn/html/202210/25/29970.html
[2] Pogrebnyakov N, Maldonado E. Didn’t roger that: social media message complexity and situational awareness of emergency responders[J]. International Journal of Information Management, 2018, 40: 166 - 174. DOI:10.1016/j.ijinfomgt.2018.02.004.
[3] Simon T, Goldberg A, Adini B. Socializing in emergencies—a review of the use of social media in emergency situations[J]. International Journal of Information Management, 2015, 35(5): 609 - 619. DOI:10.1016/j.ijinfomgt.2015.07.001.
[4] Wirz C D, Xenos M A, Brossard D, et al. Rethinking social amplification of risk: social media and zika in three languages[J]. Risk Analysis, 2018, 38(12): 2599 - 2624. DOI:10.1111/risa.13228.
[5] Zhang W, Wang M, Zhu Y. Does government information release really matter in regulating contagion-evolution of negative emotion during public emergencies? from the perspective of cognitive big data analytics[J]. International Journal of Information Management, 2020, 50: 498 - 514. DOI:10.1016/j.ijinfomgt.2019.04.001.
[6] 陈凤娇,刘德海.重大传染病多渠道疫情信息发布模式的演化路径分析[J/OL].中国管理科学:1-15[2023-02-01].https://doi.org/10.16381/j.cnki.issn1003-207x.2021.0793.
[7] 程慧平,于欢欢,蒋星.重大突发公共卫生事件中社交媒体信息过载的前因后果模型研究[J/OL].情报科学:1-12[2023-02-01].http://kns.cnki.net/kcms/detail/22.1264.G2.20221215.1717.004.html.
[8] 徐元,毛进,李纲.面向突发事件应急管理的社交媒体多模态信息分析研究[J].情报学报,2021,40(11):1150-1163.
[9] 王晰巍,王楠阿雪.数智驱动的重大突发事件应急情报管理:新机遇、新挑战、新趋势[J].图书情报工作,2022,66(16):4-12.DOI:10.13266/j.issn.0252-3116.2022.16.001.
[10] 国家突发公共事件总体应急预案[EB/OL].[2006-02-20].https://www.mem.gov.cn/xw/jyll/200602/t20060220_230269.shtml
[11] 二十大报告提出完善国家应急管理体系，应该怎么做？[EB/OL].[2022-10-17].https://baijiahao.baidu.com/s?id=1746930162528500001&wfr=spider&for=pc
[12] 查先进,张晋朝,严亚兰,李晶.网络信息行为研究现状及发展动态述评[J].中国图书馆学报,2014,40(04):100-115.DOI:10.13530/j.cnki.jlis.140017.
[13] 刘丽娟,袁曦临,杨明芳.锚定用户需求痛点的学科服务钩状模型研究[J].

图书情报工作,2019,63(15):60-67.DOI:10.13266/j.issn.0252-3116.2019.15.007.
[14] 曹锦丹,兰雪,李桂玲.基于跨理论模型的信息行为情境及其相关变量关系探讨[J].情报资料工作,2016(02):11-15.
[15] 曹锦丹,兰雪,邹男男.健康风险认知与信息交互行为关联模型研究[J].图书情报工作,2019,63(06):12-19.DOI:10.13266/j.issn.0252-3116.2019.06.002.
[16] 卢恒,张向先,吴雅威,张莉曼.基于扎根理论的学术APP用户信息交互行为影响因素研究[J].情报杂志,2021,40(04):170-178.
[17] 王晰巍,李玥琪,邱程程,毕樱瑛.直播平台用户信息交互行为图谱及特征研究[J].图书情报知识,2021,38(04):15-26+61.DOI:10.13366/j.dik.2021.04.015.
[18] 丁绪武,吴忠,夏志杰.社会化电子商务用户兴趣图谱构建的研究[J].情报理论与实践,2015,38(03):90-94.DOI:10.16353/j.cnki.1000-7490.2015.03.019.
[19] 张东鑫,张敏.图情领域LDA主题模型应用研究进展述评[J].图书情报知识,2022,39(06):143-157.DOI:10.13366/j.dik.2022.06.143.
[20] Hagen L. Content analysis of e-petitions with topic modeling: how to train and evaluate LDA models?[J]. Information Processing & Management, 2018, 54(6): 1292-1307.
[21] 张柳,王晰巍,黄博,刘婷艳.基于LDA模型的新冠肺炎疫情微博用户主题聚类图谱及主题传播路径研究[J].情报学报,2021,40(03):234-244.
[22] 霍朝光,董克,司湘云.国内外LIS学科主题热度演化分析与预测[J].图书情报知识,2021,No.200(02):35-47+57.DOI:10.13366/j.dik.2021.02.035.
[23] Sahoo S R, Gupta B B. Multiple features based approach for automatic fake news detection on social networks using deep learning[J]. Applied Soft Computing, 2021, 100: 106983.
[24] BRINS. The Page Rank citation ranking: bringing order to the web [J]. Proceedings of ASIS, 1998, 98: 161-172.
[25] 郭博,许昊迪,雷水旺.知乎平台用户影响力分析与关键意见领袖挖掘[J].图书情报工作,2018,62(20):122-132.DOI:10.13266/j.issn.0252-3116.2018.20.014.
[26] Lamberti P W, Majtey A P. Non-logarithmic jensen-shannon divergence[J]. Physica A-Statistical Mechanics and Its Applications, 2003, 329(1 - 2): 81 - 90.
[27] Zareie A, Sheikhahmadi A, Jalili M. Identification of influential users in social networks based on users’ interest[J]. Information Sciences, 2019, 493: 217-231.
[28] Li B, Feng Y, Xiong Z, et al. Research on ai security enhanced encryption algorithm of autonomous iot systems[J]. Information Sciences, 2021, 575: 379 - 398.
[29] 微博2021年营收达22.6亿美元，用户量级与活跃度均有上涨[EB/OL].[2022-03-03].https://baijiahao.baidu.com/s?id=1726274843200311246&wfr=spider&for=pc

[30] 李倩,王帅.LDA 模型下我国公共图书馆微信平台阅读推广内容主题研究[J].图书情报工作,2022,66(08):72-83.DOI:10.13266/j.issn.0252-3116.2022.08.008.
[31] 蒋甜,刘小平,刘会洲.基于关键词关联度指标(KRI)进行 LDA 噪声主题过滤的方法研究[J].图书情报工作,2020,64(03):92-99.DOI:10.13266/j.issn.0252-3116.2020.03.010.
[32] Alexa. Traffic detail (sina.com.cn)[EB/OL]. [2020-03-15]. http://www.alexa.com/siteinfo/www.sina.com.cn.
[33] 张思龙,王兰成,娄国哲.基于知识图谱的网络舆情研判系统研究[J].现代情报,2021,41(04):10-16.
[34] 曾子明,李青青,孙守强等.面向突发公共卫生事件网络舆情的事理图谱构建及演化分析[J/OL].情报理论与实践:1-13[2023-07-06].http://kns.cnki.net/kcms/detail/11.1762.G3.20230320.1524.004.html.
[35] Huang W D, Wang Q, Cao J. Tracing public opinion propagation and emotional evolution based on public emergencies in social networks[J]. International Journal of Computers Communications & Control, 2018, 13(1): 129-142.
[36] Cai M, Luo H, Meng X, et al. Network distribution and sentiment interaction: information diffusion mechanisms between social bots and human users on social media[J]. Information Processing & Management, 2023, 60(2): 103197.
[37] Qian C, Mathur N, Zakaria N H, et al. Understanding public opinions on social media for financial sentiment analysis using ai-based techniques[J]. Information Processing & Management, 2022, 59(6): 103098.
[38] Deep Prakash C, Majumdar A. Analyzing the role of national culture on content creation and user engagement on twitter: the case of indian premier league cricket franchises[J]. International Journal of Information Management, 2021, 57: 102268.
[39] Wiesenberg M, Tench R. Deep strategic mediatization: organizational leaders ’ knowledge and usage of social bots in an era of disinformation[J]. International Journal of Information Management, 2020, 51: 102042.
[40] Haouari F, Elsayed T, Mansour W. Who can verify this? finding authorities for rumor verification in twitter[J]. Information Processing & Management, 2023, 60(4): 103366.

作者贡献说明:

王晰巍：提出研究命题、研究思路及论文修订
李聪聪：负责论文撰写、修改及数据处理
孙　哲：负责数据采集
李玥琪：协助论文修改

Research on the Construction of Topic Evolution Graph for Social Group Information Behavior and Public Opinion during Significant Emergencies

Wang Xiwei [1, 2, 3], Li Congcong [1], Sun Zhe[1], Li Yueqi [1]

(1. School of Business and Management, Jilin University, Changchun, 130015)

(2. Research Center for Big Data Management, Jilin University, Changchun, 130015)

(3. Research Center for Cyberspace Governance, Jilin University, Changchun 130015)

Abstract: [Purpose/Significance] The study aims to explore the construction of a social group information behavior and public opinion topic graph during significant emergencies. This research helps to better understand the evolution of public opinion during such events and identify sensitive public opinion topics. By doing so, it enables positive public opinion guidance, which plays a crucial role in emergency response and maintaining social stability. **[Method/Process]** This paper utilizes Latent Dirichlet Allocation (LDA) to construct a graph for the clustering and evolution of public opinion topics related to social group information behavior. Additionally, it builds a graph depicting the evolution of public opinion topic popularity based on temporal characteristics. Lastly, it constructs a graph illustrating the evolution of public opinion topic paths based on similarity. Furthermore, a process model for constructing the social group information behavior and public opinion topic graph during significant emergencies is presented. **[Result/Conclusion]** The research results indicate that the proposed analysis model for public opinion topics during significant emergencies effectively analyzes the characteristics of public opinion topics related to social group information behavior on social media. It also identifies factors influencing the popularity of these topics and identifies optimal paths for topic diffusion. This study provides a new theoretical framework and analysis methods for analyzing social group information behavior and public opinion during significant emergencies. It also offers valuable insights for guiding online public opinion and public opinion governance during such events.

Keywords: Significant emergencies; social media sentiment; topic evolution; group information behavior